Anomaly based intrusion detection using ensemble machine learning and block-chain

Mekala Srinivasa Rao, Shaik Nazma, Kumbhagiri Nava Chaitanya, Thota Ambica

Department of Computer Science and Engineering, Lakireddy Bali Reddy College of Engineering, Mylavaram, India

Article Info

Article history:

Received Jan 10, 2024 Revised Feb 20, 2024 Accepted Mar 4, 2024

Keywords:

Block-chain technology Hyper-ledger fabric Intrusion detection system Machine learning Random forest

ABSTRACT

A major issue facing the quickly evolving technological world is the surge in security concerns, particularly for critical internet of things (IoT) applications like health care and the military. Early security attack detection is crucial for safeguarding important resources. Our research focuses on developing an anomaly-based intrusion detection system (IDS) using machine learning (ML) models. With the use of voting strategies, bagging ensemble, boosting ensemble, and random forest, we created a robust and long-lasting IDS. The F1 score is a crucial metric for measuring accurate predictions at the class level and serves as the focus of these ML systems. Maintaining a high F1 score in critical applications highlights the constant need for development. Make use of the latest Canadian Institute for cybersecurity internet of things (CICIoT) 2023 dataset employ hyperledger fabric to create a private channel in order to bolster the security of our IDS through the usage of block-chain technology. We use block-chain's immutable record and cryptographic techniques to establish a decentralized, tamper-proof environment. Consequently, our proposed approach provides an efficient IDS that significantly enhances resource protection and alerting the user in prior with intruder information incritical regions for IoT security applications.

This is an open access article under the CC BY-SA license.



2754

Corresponding Author:

Shaik Nazma

Department of Computer Science and Engineering, Lakireddy Bali Reddy College of Engineering

Mylavaram, Andhra Pradesh, India Email: sknazma2003@gmail.com

1. INTRODUCTION

Rapid technological innovation brings with it a serious challenge: a rising threat posed by increased security vulnerabilities. This is especially true in the internet of things (IoT), where networked gadgets play important roles in critical areas such as national security and healthcare. Cyber assaults in this setting might have serious consequences, emphasizing the fundamental need of early intrusion detection in securing important resources. In response to this requirement, our study focuses on the creation of a robust anomaly-based intrusion detection system (IDS) using machine learning (ML) [1]. Our focus is on developing an adaptive IDS using ensemble methodologies such as bagging, boosting, and random forest ensemble [2]. The F1 score, a critical indicator for exact class-level predictions, especially in high-stakes scenarios, guides the evaluation of these models. Recognizing the importance of the IoT security environment, we investigate the use of blockchain technology to strengthen our IDS. Our objective is to create a decentralized, tamper-proof ecosystem by exploiting the secure private channel provided by hyperledger fabric. The intrinsic characteristics of blockchain, such as cryptographic security and immutable record-keeping, improve the dependability of our suggested method.

Essentially, our study tackles the critical need for improved security in the linked world of IoT. We foresee an IDS that is not only more resilient but also more effective, providing enhanced protection for essential resources in crucial industries by combining robust ML models with the unprecedented security of blockchain technology. To back up our findings, we used the Canadian Institute for cybersecurity internet of things (CICIOT) 2023 dataset [3], which has a complete depiction of real-world IoT assaults. This dataset contains a variety of attack methods aimed at 105 real-world IoT devices divided into seven categories. Its inclusion of attack types [4] not often seen in other datasets, along with the collection of benign IoT traffic in both idle and active stages, makes it important for testing the efficacy of ML algorithms in constructing strong IDS. In order to put our findings into practice, we used the CICIOT-2023 dataset and thorough data preparation. Handling null values, environmentally responsible data training, and a complete set of methods spanning data purification, integration, feature selection, transformation, normalization/scaling, categorical value handling, and computational overhead reduction were all part of the process. Following that, the data is trained using ML algorithms [5], [6] and ensemble techniques such as random forest, adaptive boosting (AdaBoost), bootstrap aggregating (Bagging), and voting ensemble. The examination of performance measures comprises target classes, accuracy, recall, F1 score, and support.

The combination of ML methods and blockchain technologies to address cyber threats [7], [8] in the IoT area is a novel path worth further investigation. While blockchain has many advantages, its implementation is fraught with difficulties such as storage restrictions, scalability limitations, and vulnerabilities. Our solution offers heightened situational awareness and supports proactive efforts to neutralize and resolve any security breaches by rapidly alerting end-users of detected dangers via email notifications. This integrated alerting system is crucial in improving the overall efficacy of the IDS [9], resulting in a more responsive and adaptable security framework for protecting vital resources in IoT applications. In conclusion, our research provides a complete method to resolving the ever-changing difficulties of IoT security. We want to contribute to the creation of a more robust and effective IDS by integrating sophisticated ML models with blockchain technology, therefore bolstering the security of important resources in crucial industries.

This research expands on earlier investigations into the difficulties and weaknesses present in IoT systems. According to Narayan *et al.* [10], strong security solutions designed for IoT applications are essential. The goal of their work was to improve current IDS by integrating ML techniques and potentially ensemble methodologies to adapt to the dynamic nature of IoT landscapes. Specifically, they worked on developing an intelligent intrusion detection system (IIDS) tailored for IoT environments. Similar to this, Alsharif *et al.* [11] suggested a novel strategy for resolving IoT security issues by utilizing blockchain technology and ML techniques. Their work aims to support IDS in IoT environments by leveraging ML to improve flexibility and accuracy. Additionally, blockchain is introduced to offer a decentralized, tamper-proof framework for validating intrusion-related data. Further, Khonde and Ulagamuthalvi [12] combined blockchain technology with a hybrid IDS to present a revolutionary cybersecurity strategy. Their hybrid approach used a blockchain architecture to guarantee the integrity and immutability of data related to intrusions with several detection techniques to increase overall accuracy.

The research presented here adds to the rapidly developing subject by providing a thorough approach to deal with the constantly changing problems associated with IoT security. We want to build a more reliable and efficient IDS by fusing advanced ML models with blockchain technology [13]. This will improve the security of vital resources in important businesses. In addition to improving situational awareness, our method facilitates proactive steps to reduce security breaches, guaranteeing a more flexible and responsive security framework for IoT applications.

The CICIoT2023 dataset is a comprehensive resource for IoT security research and development. It provides real-time network traffic [10] data collected from a large network of 105 different IoT devices, simulating 33 assaults [14], [15] over seven primary categories. The categories are i) distributed denial-of-service (DDoS): flooding target systems with traffic to cause service disruption, ii) denial-of-service (DoS): depleting certain device resources to make them unavailable, iii) reconnaissance: information gathering about the network and its devices in preparation for future assaults, iv) online based: taking advantage of flaws in online apps or services operating on IoT devices, v) brute-force: repeated trial-and-error login attempts to acquire illegal access, vi) spoofing: the practice of forging source or destination addresses in order to trick other devices or systems, and vii) mirai Botnet: using vulnerabilities to infect devices and establish a botnet for coordinated attacks.

2. METHOD

2.1. A voting ensemble approach for enhancing intrusion detection

In order to improve the effectiveness of our IDS, we used a voting ensemble approach that combines the predictive capability of three well-known ML algorithms: random forest, Bagging, and AdaBoost (Figure 1). A voting ensemble is used to aggregate the individual predictions from these several algorithms

and determine the final classification based on a majority vote. This strategy tries to leverage on each algorithm's strengths while limiting its particular flaws, ultimately improving the overall robustness and accuracy of the IDS. Each algorithm, random forest, Bagging, and AdaBoost, contributes distinct benefits to the ensemble, resulting in a more complete and resilient detection method. Our ensemble technique aims to achieve superior performance in recognizing and mitigating security vulnerabilities inside IoT systems by leveraging the CICIoT-2023 dataset, which contains a broad range of real-world IoT assaults.

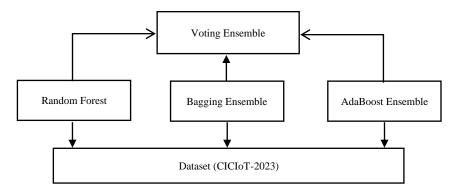


Figure 1. Voting ensemble of the input dataset

2.1.1. The voting ensemble's mechanics

The proposed voting ensemble uses the three algorithms that are random forest, bagging ensemble, and AdaBoost ensemble to determine the final classification.

- Random forest: known for its capacity to handle complicated datasets and minimize overfitting, random
 forest adds a wide variety of decision trees to the ensemble, increasing its resistance to changing data
 trends and possible outliers.
- Bagging: with an emphasis on variance reduction and better stability, Bagging delivers an ensemble of
 models trained on distinct subsets of the dataset, adding to the IDS overall generalization performance.
- AdaBoost: AdaBoost iteratively modifies the weights of misclassified examples to increase the model's capacity to recognize subtle patterns and improve detection accuracy.

Our ensemble's synergy of random forest, Bagging, and AdaBoost attempts to develop a more comprehensive and resilient detection approach. The distinct qualities of each algorithm contribute to the ensemble's combined competency, delivering a robust and diverse intrusion detection capacity. We use the CICIOT-2023 dataset to thoroughly test and optimize our proposed model. This dataset, which includes a wide range of real-world IoT attacks, provides a solid testing ground for evaluating the ensemble's effectiveness in detecting and mitigating security vulnerabilities in IoT devices. The core of the voting ensemble is its capacity to combine the many views provided by random forest, Bagging, and AdaBoost. The ensemble strategy harnesses the aggregate wisdom of these algorithms by allowing each algorithm to cast its "vote" based on its own forecast. The majority voting process enables a definite and strong final categorization in times of uncertainty or opposing forecasts. Our voting ensemble technique is a complex integration mechanism that maximizes the predictive capacity of various algorithms to strengthen the IDS's overall efficacy. This method not only highlights the virtues of random forest, Bagging, and AdaBoost, but it also highlights the promise of ensemble approaches in enhancing intrusion detection capabilities within the complex environment of IoT security.

2.1.2. Improving IoT security with blockchain integration

We use blockchain technology, specifically hyperledger fabric [16], to improve network security. Within this fabric network, we construct a channel between two organizations that are permitted to interact [17]. It is important to note that only users from these organizations can create transactions and add them to the blockchain [18]. If an intrusion occurs, the peer who detects the attack with help of ML model, is responsible for creating a transaction that includes the intruder's signature [19]. The immutability of this transaction throughout the blockchain means that all network users may quickly view the intruder's signature, essentially advertising the threat [20], [21]. An automatic alert email with the intruder's data is delivered to every network user for added awareness, boosting the security response. This multi-layered method, which

combines real-time blockchain updates with proactive email notifications, dramatically improves network security by quickly notifying all users of possible attacks and the accompanying consequences [22], [23].

2.1.3. Project workflow

The procedure begins with the ML model inspecting incoming data and discriminating between normal and abnormal network behavior. When an intrusion is detected, the system creates a transaction that contains essential facts about the assault. These transactions are arranged into a block on the hyperledger fabric blockchain, maintaining the consolidated record's integrity and transparency. Authorized entities safely propagate the blocks over the network via a private channel for transaction creation capabilities, providing real-time insights into possible risks to all network users [24] as shown in Figure 2.

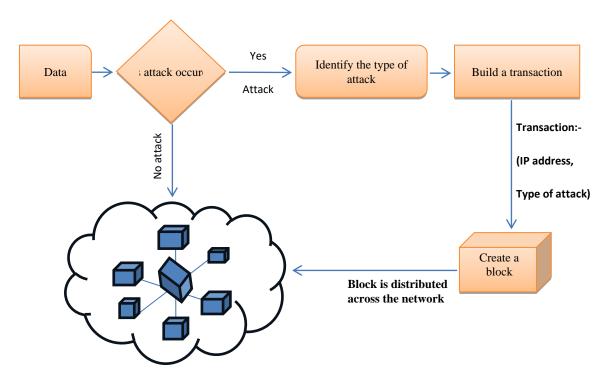


Figure 2. Workflow of proposed system

3. RESULTS AND DISCUSSION

As a key component of our IDS, the given dataset is normalized using standardization as shown in (1) then we used a hard voting ensemble technique in our project. To reach a final conclusion on the categorization of a particular instance, this approach aggregates predictions from separate models within the ensemble, namely random forest, Bagging, and AdaBoost. In hard voting, the class with the most votes from the individual models is chosen as the ultimate forecast [25]. If M represents the collection of individual models, ci represents the projected class by model i, and y represents the final prediction, the hard voting process may be stated mathematically as shown in (2). After the hard voting process the result output class performance is evaluated using precision, recall and F1 score as shown in (3)-(5). The resultant values are depicted in Figure 3.

- Data normalization (standardization):

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

Where z is standardized value, x is original value, μ is mean of the data, and σ is standard deviation of the data.

Final prediction (y)

$$y = \arg \max(\sum_{i \in M} 1(ci = c))$$
 (2)

where c is each class in classification problem and M is number of individual models.

Precision, sensitivity or true positive rate (recall), and F1 score

$$Precision = \frac{TP}{TP + FP}$$
 (3)

$$Recall = \frac{TP}{TP + FN}$$
 (4)

$$F1 - Score = \frac{{}^{2*Precision*Recall}}{{}^{Precision+Recall}}$$
 (5)

The analysis of performanse metrics of proposed model is shown in Table 1. Here we can see about the evaluation of performance of a proposed model against an existing one using standard metrics such as accuracy, precision, recall, and F1 score. The proposed model consistently outperforms the existing one across all metrics as shown in Figure 3, indicating its potential for enhancing classification accuracy and reliability. The performance metrics of used ML for each attack when intruders are present is shown in Figure 4. The generated F1 score of proposed trained ML model and existing model is shown in Figure 5.

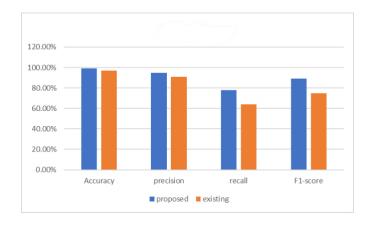
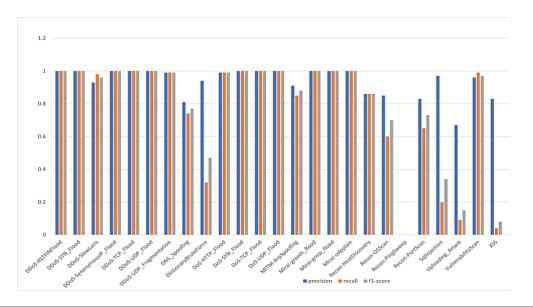


Figure 3. Comparision between existing system and proposed system

Table 1. Analysis of performance metrics

Model	Accuracy	Precision	Recall	F1 score
Random forest	98.5	91	76	79
Bagging ensemble	97	93	72	81
AdaBoost	96	80	73	75
Voting ensemble	99	95	78	89



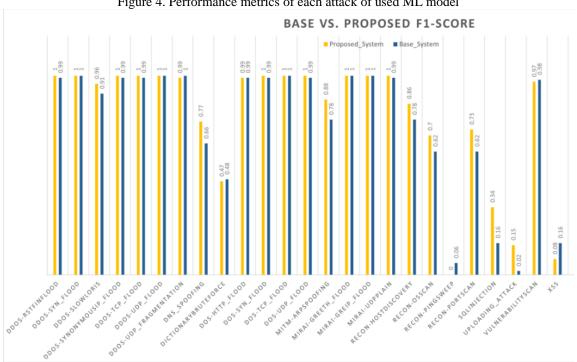


Figure 4. Performance metrics of each attack of used ML model

Figure 5. F1 score comparison for existing and proposed method

We created a static web page that extracts data from a dataset using important parameters from the supplied dataset as shown in Figure 6. The proposed trained ML model will predict the class of the attack when the intruders present and shows the intruders data as shown in Figure 7. The intrusion data is stored in the blockchain network using CouchDB as shown in Figure 8. An email will be send to the user to help the user discover suspicious behavior in the data as shown in Figure 9.





2760 ISSN: 2252-8938

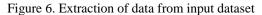


Figure 7. Prediction of intruder data

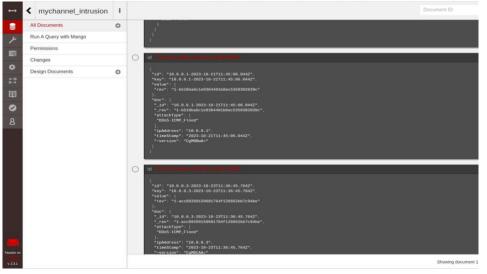


Figure 8. Storage of data in blockchain using couch database



Figure 9. Email alert send to user

CONCLUSION 4.

Finally, this study addresses the important security challenges in IoT applications, particularly in crucial industries such as health care and military. Using advanced ML models, we created an effective IDS and produced encouraging results, setting the framework for future advancements. The use of block-chain technology improves system security by creating a decentralized and tamper-proof environment. Overall, our method represents an important advancement in IoT security, with a priority on resource protection and delivering a robust solution for a growing technological context. Further more, we aim to enhance detection accuracy by integrating deep learning with ensemble ML, refine hyperparameters, and implement real-time monitoring for proactive threat response. We'll also optimize scalability, explore privacy techniques, enhance resilience against attacks, deploy IDS in edge computing, and collaborate for standardization to advance IoT security.

REFERENCES

- A. Verma and V. Ranga, "Machine learning based intrusion detection systems for IoT applications," Wireless Personal Communications, vol. 111, no. 4, pp. 2287-2310, 2020, doi: 10.1007/s11277-019-06986-8.
- D. E. Denning, "An intrusion-detection model," IEEE Transactions on Software Engineering, vol. 13, no. 2, pp. 222-232, 1987, doi: 10.1109/TSE.1987.232894.
- E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: a real-time dataset and
- benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, pp. 1–26, 2023, doi: 10.3390/s23135941. M. Anwer, M. Umer, S. M. Khan, and Waseemullah, "Attack detection in IoT using machine learning," *Engineering, Technology* and Applied Science Research, vol. 11, no. 3, pp. 7273–7278, 2021, doi: 10.48084/etasr.4202.
- P. Ioulianou, V. Vasilakis, I. Moscholios, and M. Logothetis, "A signature-based intrusion detection system for the internet of things," in Smart Cities Symposium 2018, 2018, pp. 1-6.
- M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, vol. 40, no. 5, pp. 516-524, 2010, doi: 10.1109/TSMCC.2010.2048428
- R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, G. Sahil, and M. M. Hassan, "A distributed intrusion detection system to detect DDoS attacks in blockchain-enabled IoT network," Journal of Parallel and Distributed Computing, vol. 164, pp. 55-68, 2022, doi: 10.1016/j.jpdc.2022.01.030.
- F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in IoT security: current solutions and future challenges," IEEE Communications Surveys and Tutorials, vol. 22, no. 3, pp. 1686-1721, 2020, doi: 10.1109/COMST.2020.2986444.
- M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," Journal of Network and Computer Applications, vol. 60, pp. 19-31, 2016, doi: 10.1016/j.jnca.2015.11.016.
- [10] K. R. Narayan, S. Mookherji, V. Odelu, R. Prasath, A. C. Turlapaty, and A. K. Das, "IIDS: design of intelligent intrusion detection system for internet-of-things applications," in 2023 IEEE 7th Conference on Information and Communication Technology (CICT), 2023, pp. 1–6, doi: 10.1109/CICT59886.2023.10455720.
- N. A. Alsharif, S. Mishra, and M. Alshehri, "IDS in IoT using Machine Learning and Blockchain," Engineering, Technology and Applied Science Research, vol. 13, no. 4, pp. 11197–11203, 2023, doi: 10.48084/etasr.5992.

 S. R. Khonde and V. Ulagamuthalvi, "Hybrid intrusion detection system using blockchain framework," Eurasip Journal on
- Wireless Communications and Networking, vol. 2022, no. 1, pp. 1-25, 2022, doi: 10.1186/s13638-022-02089-4.
- B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. D. Alvarenga, "A survey of intrusion detection in internet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, 2017, doi: 10.1016/j.jnca.2017.02.009.
- S. Raza, L. Wallgren, and T. Voigt, "SVELTE: real-time intrusion detection in the internet of things," Ad Hoc Networks, vol. 11, no. 8, pp. 2661–2674, 2013, doi: 10.1016/j.adhoc.2013.04.014.
- M. Ammar, G. Russello, and B. Crispo, "Internet of things: a survey on the security of IoT frameworks," Journal of Information Security and Applications, vol. 38, pp. 8-27, 2018, doi: 10.1016/j.jisa.2017.11.002.
- Linux Foundation, "Hyperledger blockchain for business," Hyperledger. Accessed: Oct. 01, 2017. [Online]. Available: https://www.hyperledger.org/
- S. Aggarwal and N. Kumar, "Core components of blockchain," Advances in Computers, vol. 121, pp. 193-209, 2021, doi: 10.1016/bs.adcom.2020.08.010.
- [18] A. Ramachandran and D. M. Kantarcioglu, "Using blockchain and smart contracts for secure data provenance management," Arxiv-Computer Science, 2017.
- W. Meng, E. W. Tischhauser, Q. Wang, Y. Wang, and J. Han, "When intrusion detection meets blockchain technology: a review," IEEE Access, vol. 6, pp. 10179-10188, 2018, doi: 10.1109/ACCESS.2018.2799854.
- M. Abomhara and G. M. Køien, "Cyber security and the internet of things: vulnerabilities threats intruders and attacks," Journal of Cyber Security and Mobility, vol. 4, no. 1, pp. 65-88, 2015, doi: 10.13052/jcsm2245-1439.414.
- X. Wang et al., "Survey on blockchain for Internet of Things," Computer Communications, vol. 136, pp. 10-29, 2019, doi: 10.1016/j.comcom.2019.01.006.
- E. S. Babu et al., "Blockchain-based intrusion detection system of IoT urban data with device authentication against DDoS attacks," Computers and Electrical Engineering, vol. 103, 2022, doi: 10.1016/j.compeleceng.2022.108287.
- E. S. Babu, M. S. Rao, R. Pemula, S. R. Nayak, and A. Shankar, "A hybrid intrusion detection system against botnet attack in IoT using light weight signature and ensemble learning technique," Research Square, vol. 1, pp. 1-17, 2022, doi: 10.21203/rs.3.rs-905197/v1
- M. A. Aydin, A. H. Zaim, and K. G. Ceylan, "A hybrid intrusion detection system design for computer network security," Computers and Electrical Engineering, vol. 35, no. 3, pp. 517-526, 2009, doi: 10.1016/j.compeleceng.2008.12.005.
- S. Emanet, G. K. Baydogmus, and O. Demir, "An ensemble learning based IDS using voting rule: VEL-IDS," PeerJ Computer Science, vol. 9, pp. 1-23, 2023, doi: 10.7717/PEERJ-CS.1553.

BIOGRAPHIES OF AUTHORS





Shaik Nazma currently pursuing her bachelor's degree in the field of Computer Science and Engineering, at Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India. She has done many projects as a part of her B.Tech. curriculum. She has contributed her best in her academics. She can be contacted at email: sknazma2003@gmail.com.





Thota Ambica currently pursuing her bachelor's degree in the field of Computer Science and Engineering, at Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India. She has done many projects as a part of her B.Tech. curriculum. She has contributed her best in her academics. She can be contacted at email: ambicat2003@gmail.com.